

534,764
10/534764
(12) NACH DEM VEREIN ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
27. Mai 2004 (27.05.2004)

PCT

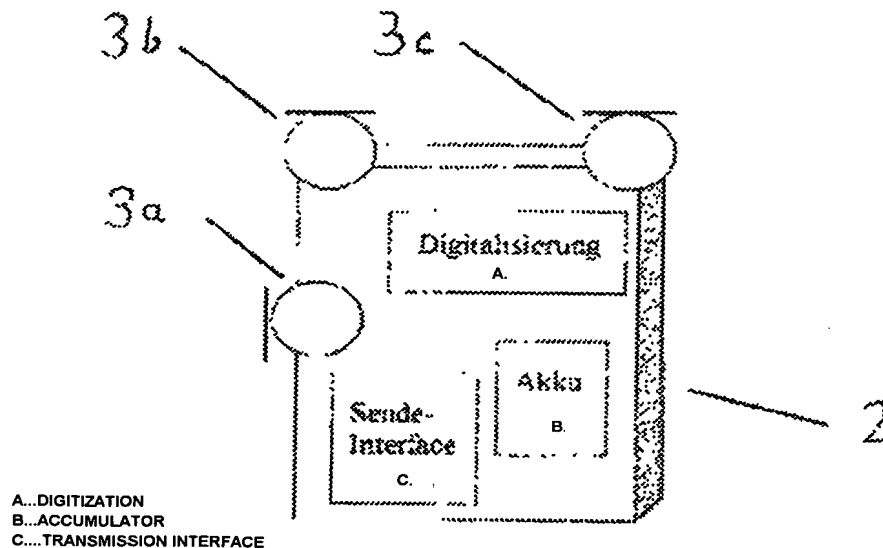
(10) Internationale Veröffentlichungsnummer
WO 2004/044889 A2

- (51) Internationale Patentklassifikation⁷: G10L 15/22
- (21) Internationales Aktenzeichen: PCT/EP2003/012660
- (22) Internationales Anmeldedatum:
12. November 2003 (12.11.2003)
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität:
102 52 457 12. November 2002 (12.11.2002) DE
- (71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von
US): **HARMAN BECKER AUTOMOTIVE SYSTEMS
GMBH** [DE/DE]; Becker-Göring-Strasse 16, 78126 Karls-
bad (DE).
- (72) Erfinder; und
- (75) Erfinder/Anmelder (nur für US): **BIRK, Christian**
[DE/DE]; Irisweg 33, 89079 Ulm (DE). **HAULICK, Tim**
[DE/DE]; In den Schlosswiessen 14, 89143 Blaubeuren
(DE). **LINHARD, Klaus** [DE/DE]; Gundershofen 86,
89601 Schelklingen (DE).
- (74) Anwalt: **KUHL, Dietmar**; Grünecker, Kinkeldey,
Stockmair & Schwanhäuser, Anwaltssozietät, Maxim-
ilianstrasse 58, 80538 München (DE).
- (81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH,
PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[Fortsetzung auf der nächsten Seite]

(54) Title: VOICE INPUT INTERFACE

(54) Bezeichnung: SPRACHEINGABE-INTERFACE



(57) Abstract: The invention concerns a voice input system comprising a fixed central unit (1) and a mobile voice interface (2) carried by a user, for example in his clothes. The voice interface (2) is used for sound input of voice commands which are transmitted after being electronically processed by radio relay channel to the central unit (1). The invention is characterized in that the voice interface comprises at least two microphones (3a, 3b, 3c) which interact in the form of a microphone array, thereby enabling part of a directional characteristic to be obtained during the sound input and the noise to be greatly reduced. The central unit (1) comprises the elements required for voice recognition, for example a processor system, and interfaces (4a, 4b, 4c) for controlling external appliances through radio relay channel or via cable.

[Fortsetzung auf der nächsten Seite]

WO 2004/044889 A2



(84) **Bestimmungsstaaten (regional):** ARIPO-Patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(57) **Zusammenfassung:** Beschrieben wird ein Spracheingabesystem, bestehend aus einer stationären Zentraleinheit (1) und einem transportablen, vom Anwender z.B. an der Kleidung mitgeführten Sprachinterface (2). Das Sprachinterface (2) dient zur akustischen Aufnahme gesprochener Kommandos, die nach elektronischer Aufbereitung drahtlos an die Zentraleinheit (1) übertragen werden. Dabei enthält das Sprachinterface zwei oder mehr Mikrofone (3a, 3b, 3c), die als Mikrofonarray zusammenwirken, wodurch einerseits eine Richtcharakteristik bei der Schallaufnahme und andererseits eine weitgehende Unterdrückung von Störgeräuschen ermöglicht wird. Die Zentraleinheit (1) enthält die Komponenten der Spracherkennung, z.B. Prozessorsystem, und Schnittstellen (4a, 4b, 4c), über die externe Geräte drahtlos oder über Kabel gesteuert werden.

Spracheingabe-Interface

Die Erfindung betrifft Systeme, bei denen das Sprechen eines Anwenders aufgenommen und als Signal weitergeleitet wird.

- 5 Insbesondere betrifft die Erfindung Systeme, bei denen über eine Spracheingabe Funktionen ausgelöst oder gesteuert werden.

Systeme mit Spracheingabe und Sprachsteuerung sind bekannt.

- 10 So sind beispielsweise bei Computern Eingaben - Diktieren von Texten oder Steuerbefehle - per Sprache möglich. In der Medizintechnik existieren Geräte, die zur Entlastung bei schwierigen Eingriffen durch Sprachkommandos des Arztes gesteuert werden können. Auch im Sicherheitsbereich werden Vorrichtungen
15 eingesetzt, die beispielsweise eine verschlossene Tür selektiv nur auf Spracheingaben autorisierter Personen öffnen.

Alle derartigen System erfordern für das Erkennen der sprachlichen Information bzw. Kommandos eine hohe akustische Qualität der Spracheingabe. Besonders störend wirkt sich beispielsweise ein zu großer Abstand zwischen Mikrofon und Mund des Sprechers aus (zu geringe Eingabelautstärke). In gleicher Weise wirkt ein Sprechen in eine Richtung, die nicht im
20 Hauptaufnahmebereich des Mikrofons liegt. Auch eine relativ
25 nahe Distanz zum Mikrofon kann sich negativ auswirken, da in diesem Fall zum einen leicht eine Übersteuerung des Aufnahmepegels auftreten kann, zum anderen durch den Atmen des Sprechers zusätzlich starke akustische Störgeräusche (Windgeräusche) entstehen können. Grundsätzlich wirkt sich auch ein ho-

her Geräuschpegel aus der Umgebung sehr störend auf die Erkennungsgenauigkeit des Spracheingabesystems aus.

Um diese bekannten Probleme zu lösen, werden beispielsweise
5 für Computersysteme mit Texteingabe Mikrofonbügel eingesetzt,
die (zumeist in Verbindung mit einem Kopfhörer oder Ohrhörer)
vom Anwender am Kopf getragen werden. Bei entsprechender Aus-
richtung des Bügels befindet sich das Aufnahmemikrofon auch
bei Kopfbewegungen stets in gleichem Abstand nahe am Mund des
10 Sprechers. Nachteilig ist dabei die eingeschränkte Bewegungs-
freiheit aufgrund der üblichen Kabelverbindung zum Computer.
Auch können Störgeräusche durch Bewegungen des Kabels auftre-
ten. Zudem empfinden viele Personen das Tragen von Kopf- oder
Ohrhörern, insbesondere bei längerer Anwendung, als unange-
15 nehm.

Alternativ werden daher auch Mikrofone mit stationärer Posi-
tion, z.B. Tischmikrofon mit Stativ oder integriert in das
Gehäuse (PC, Laptop) oder befestigt am Gerät (Türrahmen bei
20 Sicherungsfunktion) verwendet. Nachteilig ist dabei die auf
einen bestimmten Raumbereich vor dem Mikrofon beschränkte
Aufnahmezone. Dies erfordert vom Anwender die Einhaltung ei-
ner definierte Position, Körperhaltung, Sprechrichtung usw.,
d.h. es besteht praktisch keine Bewegungsfreiheit während der
25 Spracheingabe.

Ausgehend von diesem Stand der Technik liegt der vorliegenden
Erfindung die Aufgabe zugrunde, ein verbessertes System zur
Spracheingabe zu entwickeln, das die genannten Nachteile
30 weitgehend überwindet und zusätzliche Vorteile aufweist.

Diese Aufgabe wird bei einer Vorrichtung mit den Merkmalen
des Oberbegriffs des Anspruchs 1 gelöst durch die kennzeich-
nenden Merkmale des Anspruchs 1. Weitere Einzelheiten der Er-

findung und Vorzüge verschiedener Ausführungsformen sind Gegenstand der Merkmale der Unteransprüche.

Die erfindungsgemäße Vorrichtung wird im folgenden anhand einer bevorzugten Ausführungsform beschrieben, wobei Bezug genommen wird auf die Abbildungen und den darin aufgeführten Bezugsziffern.

Dabei zeigen:

10

Fig. 1 das erfindungsgemäße Spracheingabesystem aus Zentraleinheit und separaten Sprachinterface

15

Fig. 2 eine schematische Darstellung einer Ausführungsform des Sprachinterfaces

Fig. 3 Richtcharakteristik des vom Anwender mitgeführten Sprachinterfaces

20

Die vorliegende Erfindung zielt darauf ab, Komfort und Qualität bei Spracheingabesystemen deutlich zu verbessern. Dabei soll ein mit Mikrofonen ausgestattetes transportables Sprachinterface vom Anwender ständig getragen werden, wodurch ihm letztlich ein universeller Sprachzugang zu verschiedenen Systemen ermöglicht wird. Durch Verwendung von Mikrofonarrays kann auch bei verschiedenen akustischen Umgebungen eine hohe Eingabequalität gegenüber Störgeräuschen erreicht werden. Damit ist es auch als Spracheingabesystem in Fahrzeugen geeignet, da Störungen durch Fahrgeräusche oder Echoeffekte bei Lautsprecherausgaben vom Mikrofonarray vermindert werden. Für ein ständig zu tragenden Sprachinterfaces ist es wichtig, dass es klein und leicht ist und - je nach äußerer Gestaltung - beispielsweise als Schmuck oder Identifikationssymbol akzeptiert wird.

30

Fig. 1 zeigt als Übersicht die zusammenwirkenden Komponenten des Spracheingabesystems. Das Sprachinterface (2) ist als mobile Einheit ausgeführt und wird vom Anwender, z.B. an der Kleidung, getragen. Es überträgt die akustisch aufgenommenen Sprachsignale drahtlos, z.B. per Infrarot- oder Funkverbindung an die Zentraleinheit (1), in der die Signale weiter aufbereitet werden und diverse Steuerfunktionen ausgelöst werden.

Zur Gewährleistung einer hohen Qualität der Sprachaufnahme weist das Sprachinterface (2) zwei oder mehrere Mikrofone (3a, 3b, 3c) auf. Eine solche Anordnung ist in Fig. 2 vergrößert dargestellt.

Die verwendeten Mikrofone (3a, 3b, 3c) können individuelle Richtcharakteristiken (Niere, Hypernieren, Acht) aufweisen. Damit wird bereits durch eine vordefinierte Mikrofon-Richtcharakteristik Schall vorrangig aus einer bestimmten Zone verstärkt aufgenommen.

Die hier erfindungsgemäß vorgeschlagene Verwendung eines kleinen Mikrofonsystems mit zwei oder mehr Mikrofonen erlaubt die Bildung von Mikrofonarrays. Durch das Zusammenwirken der Mikrofone in einem solchen Mikrofonarray kann - in Verbindung mit einer für solche Arrays üblichen elektronischen Aufbereitung - die Qualität der Spracheingabe wesentlich verbessert werden: So kann - über die oben erwähnte, unveränderliche Mikrofon-Richtcharakteristik hinaus - eine spezielle räumliche Richtwirkung des Mikrofonarrays eingestellt werden, d.h. akustische Signale werden vorrangig aus einem gewünschten Raumbereich (Mundpartie des Anwenders) aufgenommen. Durch diese zusätzliche Array-Richtcharakteristik werden Nebengeräusche aus anderen Bereichen der Umgebung weiter unterdrückt bzw. können elektronisch weitgehend herausgefiltert werden.

Die Array-Richtcharakteristik ergibt sich aus der Anzahl und geometrischer Anordnung der Mikrofone. Im einfachsten Fall werden zwei Mikrofone verwendet (Minimalkonfiguration). Vorzugsweise ist das Interface jedoch mit drei (wie in Fig. 2 dargestellt) oder mehr Mikrofonen ausgestattet, die eine bessere Richtwirkung und Störschallunterdrückung ermöglichen. Es gibt zwei grundsätzliche Mikrofon-Array-Anordnungen: `broad-side` und `end-fire`. Bei `broad-side` liegt die Richtwirkung senkrecht auf der gedachten Verbindungslinie der Mikrofone, bei `end-fire` liegt die Richtwirkung in Richtung der gedachten Verbindungslinie der Mikrofone. Das Ausgangssignal eines `broad-side` Arrays ergibt sich in einfachster Form aus der Summation der einzelnen Signale, beim `end-fire` Array aus der Differenz, wobei auch Laufzeitkorrekturen eingesetzt werden.

Die Richtwirkung des Mikrofonarrays kann durch weitere Maßnahmen verändert und damit eine adaptive Richtcharakteristik erreicht werden. Dabei werden die einzelnen Mikrofonsignale nicht einfach addiert oder subtrahiert, sondern durch spezielle Signalverarbeitungsalgorithmen so ausgewertet, dass akustische Signale aus einer Hauptrichtung verstärkt empfangen werden und Nebengeräusche aus anderen Richtungen abgeschwächt aufgenommen werden. Die Lage der Hauptrichtung ist dabei veränderbar, d.h. sie kann adaptiv an ein sich änderndes akustisches Szenario angepaßt werden. Dabei kann beispielsweise die Art und Weise, wie das Signal der Hauptrichtung bewertet und maximiert wird und gleichzeitig Störgeräusche aus Nebenrichtungen minimiert werden, in einem Fehlerkriterium festgelegt sein. Algorithmen zur Erzeugung einer adaptiven Richtwirkung sind unter dem Begriff `Beamforming` bekannt. Verbreitete Algorithmen sind z.B. der Jim-Griffith Beamformer oder der `Frost` Beamformer.

Durch Veränderung entsprechender Parameter kann die Haupt-
richtung beim `Beamforming` so variiert werden, dass sie mit
der Richtung aus der gesprochen wird übereinstimmt, was einer
aktiven Sprecherortung entspricht. Eine einfache Möglichkeit
5 die Sprechrichtung zu bestimmen ist z.B. die Laufzeitschät-
zung zwischen zwei Empfangssignalen zweier Mikrofone. Wird
die Kreuzkorrelation zwischen den beiden Werten errechnet, so
ergibt sich der maximale Kreuzkorrelations-Wert bei der Lauf-
zeitverschiebung beider Signale. Wird nun die diese Laufzeit
10 zur Verzögerung des einen Signals eingesetzt, so sind die
beiden Signale wieder gleichphasig. Damit ist die Hauptrich-
tung in die aktuelle Sprechrichtung eingestellt. Wird die
Laufzeitschätzung und Korrektur wiederholt durchgeführt, so
wird einer Relativbewegung des Sprechers adaptiv gefolgt. Da-
15 bei ist es vorteilhaft für die Sprecherortung nur einen zuvor
festgelegten Raumsektor zuzulassen. Dazu ist es erforderlich,
die Mikrofonanordnung grob in einer gewissen Richtung zum
Mund des Sprechers anzubringen, z.B. am Kleidungsstück des
Sprechers.

20

Der Mund des Anwenders kann sich dann innerhalb des festge-
legten Raumsektors gegenüber der Position des Sprachinter-
faces (2) frei bewegen, die Sprecherortung wird ihm folgen.
Wird eine Signalquelle außerhalb des festgelegten Raumsektors
25 detektiert, dann wird diese als Störung identifiziert (bei-
spielsweise eine Lautsprecherausgabe). Der Beamforming-Algo-
rithmus kann sich nun speziell auf den Schall aus dieser
Richtung im Sinne eine Minimierung des Störsignals einstel-
len. Dies läßt auch eine wirksame Echokompensation zu.

30

Fig. 3 zeigt beispielhaft ein kleines Mikrofonsystem aus zwei
einzelnen Mikrofonen mit einer eingepprägten Richtcharakteris-
tik die rechts am Mund des Sprechers vorbei gerichtet ist.
Die Mikrofone sind hier an der oberen Kante eines kleinen Ge-
35 häuses angebracht. Der Array-Typ ist `broad-side`, d.h. die

Richtwirkung des Arrays ist senkrecht zur Kante des Gehäuses nach oben gerichtet. Die adaptive Richtwirkung über Beamforming-Algorithmen sorgt dafür, dass die wirksame Richtcharakteristik auf die Quelle, den Mund des Anwenders ausgerichtet ist.

Mikrofone mit hoher Qualität sind als Miniaturausführung (bis herab zu Millimetergröße) verfügbar. Ebenso sind drahtlose Übertragungseinrichtungen, z.B. Infrarot- oder Funksender, mit heutiger Technik (z.B. als SMD oder IC) extrem klein herstellbar. Zur Stromversorgung genügt eine kleine Batterie bzw. Akkumulator (z.B. Knopfzelle), da nur ein geringer Energieverbrauch entsteht. Damit ist die Integration aller Komponenten des Sprachinterfaces (2) zu einer kleinen und auch aufgrund des sehr geringen Gewichts bequem tragbaren Einheit möglich. Beispielsweise kann ein solches miniaturisiertes Sprachinterface (2) als Ansteckvorrichtung bzw. Clip an der Kleidung befestigt werden (vergleichbar mit einer Brosche) oder an einem Armband oder einer Halskette mitgeführt werden.

In einer ersten Ausführungsform werden die jeweiligen Signale der einzelnen Mikrofone des Arrays parallel zur Zentraleinheit übermittelt. Dort werden dann diese Signale elektronisch weiter verarbeitet, um Richtungscharakteristik und Störgeräuschunterdrückung einzustellen. Diese Funktionen können alternativ - zumindest teilweise - auch bereits im Sprachinterface selbst erfolgen. In dem Sprachinterface (2) sind dann entsprechende elektronische Schaltkreise integriert, die für eine erste Signalaufbereitung sorgen. So kann beispielsweise auch der jeweilige Mikrofon-Aufnahmepegel durch automatische Verstärkungsregelungen (Automatic-Gain-Control) angepaßt werden oder durch entsprechende Filter bestimmte Frequenzanteile abgeschwächt oder verstärkt werden.

Wie in Figur 2 dargestellt kann das Eingabeinterface (2) auch Komponenten für eine Umwandlung analoger in digitale Signale aufweisen. Die Übertragung digitaler Signale zur Zentraleinheit (1) bietet mit den heute verfügbaren Sende/Empfangstechniken einen sehr hohen Informationsfluss bei minimaler Störanfälligkeit.

Das Sprachinterface (2) kann mit einer Aktivierungstaste oder Sensorfläche ausgestattet werden, deren Betätigung/Berührung eine Spracherkennung aktiviert, beispielsweise durch Übertragung eines entsprechenden Signals an die Zentraleinheit (1). Alternativ kann eine solche Aktivierung auch durch Spracheingabe eines Schlüsselwortes erfolgen.

In einer erweiterten Ausführungsform weist das Sprachinterface (2) zusätzlich eine Empfangseinrichtung auf (nicht dargestellt), die drahtlos übermittelte Steuersignale von der Zentraleinheit (1) empfangen kann. Damit kann die Charakteristik der Sprachaufnahme (Kombination der Mikrofone zum Array, Verstärkung, Frequenzfilter usw.) im Sprachinterface (2) durch die Zentraleinheit (1) verändert werden. Weiterhin kann die Zentraleinheit (1) über die Empfangseinrichtung dem Benutzer auch ein akustisches oder optisches Signal übermitteln, beispielsweise als `Feedback` für erfolgreich ausgelöste Aktionen aber auch Fehlfunktionen (Kommando nicht erkannt oder - z.B. wegen Störung der Schnittstellen - nicht ausführbare Funktion) oder zur Anforderung weiterer Eingaben (Kommando unvollständig). Als Signalgeber können z.B. verschiedenfarbige LED's oder piezokeramische Kleinlautsprecher in das Spracheingabeinterface (2) integriert werden. Auch ist es denkbar, eine Eingabeeinheit, beispielsweise Tastatur, zur Eingabe von Informationen in Textform vorzusehen.

Die Zentraleinheit (1) empfängt die drahtlos übermittelten Signale des Sprachinterfaces (2) und wertet sie für eine

Spracherkennung aus. Auch die verschiedenen erforderlichen Kalkulationen zur Sprecherortung und adaptiven Anpassung der Richtungscharakteristik des Mikrofonarrays können ganz oder teilweise in der Zentraleinheit (1) erfolgen. Die für eine
5 zuverlässige und schnelle Signalverarbeitung und Spracherkennung erforderliche Prozessorleistung kann beispielsweise über ein übliches Computer- bzw. Mikroprozessorsystem zur Verfügung gestellt werden, das vorzugsweise Teil der Zentraleinheit (1) ist. Damit können anwenderspezifische Konfigurationen (Arraycharakteristik, Spracheigenheiten, spezielle
10 Schalt/Steuerkommandos usw.) jederzeit neu einprogrammiert werden.

Da die Zentraleinheit (1) getrennt vom Sprachinterface (2) -
15 z.B. integriert in Geräte (TV, Telefonanlage, PC-System usw.) - stationär ausgebildet ist, entstehen auch bei leistungsfähiger elektronischer und EDV-technischer Ausstattung keine besonderen Probleme oder Einschränkungen bzgl. Stromversorgung, Volumen, Gewicht oder auch Kühlung wärmeproduzierender
20 Komponenten (Prozessor).

Die Zentraleinheit (1) ist so konfiguriert, dass erkannte Kommandos über integrierte Schnittstellen (4a, 4b, 4c) entsprechende Schalt- und Steuerfunktionen auslösen, auf die
25 verschiedene externe Geräte reagieren.

Hierzu können beispielsweise Telefonsysteme, Audio- und Videogeräte gehören, aber auch eine Vielzahl elektrisch/elektronisch steuerbarer Hausgeräte (Licht, Heizung, Klimaanlage, Jalousien, Türöffner u.v.a.m). Bei Einsatz in einem Fahrzeug können entsprechend Funktionen des Fahrzeugs (Navigationssystem, Musikanlage, Klima, Scheinwerfer, Scheibenwischer usw.)
30 gesteuert werden.

Die Übertragung der Steuersignale der Schnittstellen (4a, 4b, 4c) zu den verschiedenen externen Geräten kann ihrerseits wieder drahtlos (IR, Funk) oder auch (z.B. bei Einsatz im Fahrzeug) mittels Kabelverbindung erfolgen.

5

Die erfindungsgemäße Spracheingabevorrichtung bietet dem Anwender völlige Bewegungsfreiheit (auch über größere Distanz z.B. verschiedene Räume/Etagen eines Hauses). Hinzu kommt ein optimaler Tragekomfort, da das Sprachinterface (2) nicht am Kopf befestigt werden muß. Durch das kleine Eigengewicht und die geringe Größe (insbesondere als Miniaturausführung) entsteht durch das Sprachinterface keinerlei Behinderung, so dass es - vergleichbar zu einer Armbanduhr - für ein dauerhaftes Tragen geeignet ist.

15

Verschiedene Geräte bzw. Funktionen an unterschiedlichen Orten können mit ein und demselben Sprachinterface gesteuert werden (Multimediaeinrichtungen, Hausgeräte, Fahrzeugfunktionen, Büro-PC, Sicherheitsanwendungen, usw.), da auch mehrere Zentraleinheiten an unterschiedlichen Standorten (z.B. Privaträume, Fahrzeug, Arbeitsbereich) so eingestellt werden können, dass sie auf das gleiche Sprachinterface bzw. den gleichen Anwender reagieren.

25

Besonders vorteilhaft ist die hohe Flexibilität des vorgestellten Spracheingabesystems. So ermöglicht das Mikrofonarray eine aktive - je nach Ausführung auch automatische - Justierung auf veränderte Gegebenheiten, wie z.B. andere Raumakustik oder Befestigung des Sprachinterfaces an anderer Stelle der Kleidung (andere Sprechrichtung) oder auch ein neuer Anwender (eigene Sprechweise). Durch Programmierung der Zentraleinheit kann das Spracheingabesystem flexibel an die jeweiligen Steuer und Regelungsaufgaben angepaßt werden.

30

Grundsätzlich bietet das erfindungsgemäße System den Vorteil, dass das sehr kleine und leichte Sprachinterface vom Anwender ständig getragen werden kann. Ein Benutzer kann es ohne Behinderung auch bei sich führen, wenn er seinen Aufenthaltsort
5 ändert, z.B. in sein Auto einsteigt. Durch das Mikrofonarray und damit verbundenen Richtungswirkung der Schallaufnahme ergibt sich auch bei ungünstigen Umständen (Umgebungsgeräusche) eine hohe akustische Sprachaufnahmequalität. Die aufwändigen Prozesse der Sprachaufbereitung können über entsprechend
10 leistungsfähige Komponenten in der Zentraleinheit mit sehr hoher Erkennungssicherheit erfolgen. Dabei kann die Spracherkennung individuell auf einen oder mehrere Anwender justiert werden, z.B. für eine Funktionsbeschränkung nur auf vorbestimmte, autorisierte Person(en). Die Ansteuerung verschiede-
15 ner Geräte über entsprechende Schnittstellen ist frei konfigurierbar und kann nach individuellen Wünschen für verschiedenste Anwendungszwecke modifiziert und erweitert werden.

Patentansprüche

- 5 1. Spracheingabesystem aus stationärer Zentraleinheit (1)
und transportablem Sprachinterface (2), wobei
die Zentraleinheit (1) Mittel zum Empfang drahtlos über-
mittelter Sprachsignale und Mittel für eine Auswertung
dieser Sprachsignale zur Erkennung gesprochener Worte
10 aufweist, sowie Schnittstellen (4a, 4b, 4c), die in Ab-
hängigkeit von erkannten Worten Schalt- bzw. Regelsignale
erzeugen und an externe Geräte übertragen und wobei
das Sprachinterface (2) zwei oder mehr Mikrofone (3a, 3b,
3c) aufweist, die als Mikrofonarray kombiniert sind, so-
15 wie Mittel zur drahtlosen Übertragung der vom Mikrofonar-
ray generierten Signale an die Zentraleinheit (1).
2. Spracheingabesystem nach Anspruch 1,
d a d u r c h g e k e n n z e i c h n e t ,
20 dass das Sprachinterface (2) Mittel zur elektronischen
Aufbereitung der von den Mikrofonen (3a, 3b, 3c) des Mik-
rofonarrays erzeugten Signale aufweist.
3. Spracheingabesystem nach Anspruch 2,
25 d a d u r c h g e k e n n z e i c h n e t ,
dass die elektronische Aufbereitung eine Veränderung der
jeweiligen Signalpegel (Verstärkung, Dämpfung, AGC) und/
oder Frequenzanteile (Höhen-, Tiefen-, Band-Filter) und/
oder der jeweiligen Signalphasen bewirkt.

4. Spracheingabesystem nach einem der Ansprüche 2 oder 3,
d a d u r c h g e k e n n z e i c h n e t ,
dass die Mittel zur elektronischen Aufbereitung auch Mit-
5 tel zur Umwandlung analoger in digitale Signale beinhalten.
5. Spracheingabesystem nach einem der Ansprüche 1 bis 4,
d a d u r c h g e k e n n z e i c h n e t ,
10 dass das Sprachinterface (2) Mittel zum Empfang drahtlos
(z.B. IR, Funk) von der Zentraleinheit (1) übermittelter
Signale aufweist.
6. Spracheingabesystem nach Anspruch 2, 3 oder 4,
15 d a d u r c h g e k e n n z e i c h n e t ,
dass das Sprachinterface (2) Mittel zum Empfang drahtlos
(z.B. IR, Funk) von der Zentraleinheit (1) übermittelter
Steuersignale aufweist, mit denen die Zentraleinheit (1)
die Mittel zur elektronischen Aufbereitung des Sprachin-
20 terfaces (2) beeinflussen kann.
7. Spracheingabesystem nach einem der Ansprüche 1 bis 6,
d a d u r c h g e k e n n z e i c h n e t ,
dass die Mittel der Zentraleinheit (1) zur Auswertung der
25 Sprachsignale als Prozessorsystem (EDV) ausgeführt sind.
8. Spracheingabesystem nach Anspruch 7,
d a d u r c h g e k e n n z e i c h n e t ,
dass das Prozessorsystem Anwender-bezogene Daten (z.B.
30 individuelle Sprachkommandos, Sprachparameter) speichert
und zur Auswertung der Sprachsignale und/oder zur Erzeugung
individuell zugeordneter Steuersignale einbezieht.

9. Spracheingabesystem nach einem der Ansprüche 1 bis 8,
dadurch gekennzeichnet,
dass das Mikrofon-Array (3a, 3b, 3c) eine Richtcharakteristik in der Weise aufweist, dass Schall vorrangig aus
5 einer Richtung aufgenommen wird.
10. Spracheingabesystem nach Anspruch 9,
dadurch gekennzeichnet,
dass die Richtcharakteristik durch Aufbereitung der von
10 den Mikrofonen (3a, 3b, 3c) des Mikrofonarrays erzeugten
Signale über Beamforming-Algorithmen (z.B. Griffith-,
Frost-Beamformer) verändert werden kann.
11. Spracheingabesystem nach Anspruch 10,
15 dadurch gekennzeichnet,
dass die Richtcharakteristik in der Weise verändert wird,
dass sie einer räumlichen Verlagerung der Schallquelle
folgt (Sprechernachführung).
- 20 12. Spracheingabesystem nach einem der Ansprüche 1 bis 11
dadurch gekennzeichnet,
dass die einzelnen Mikrofone (3a, 3b, 3c) des Arrays eine
Richtcharakteristik (z.B. Niere, Acht) aufweisen.

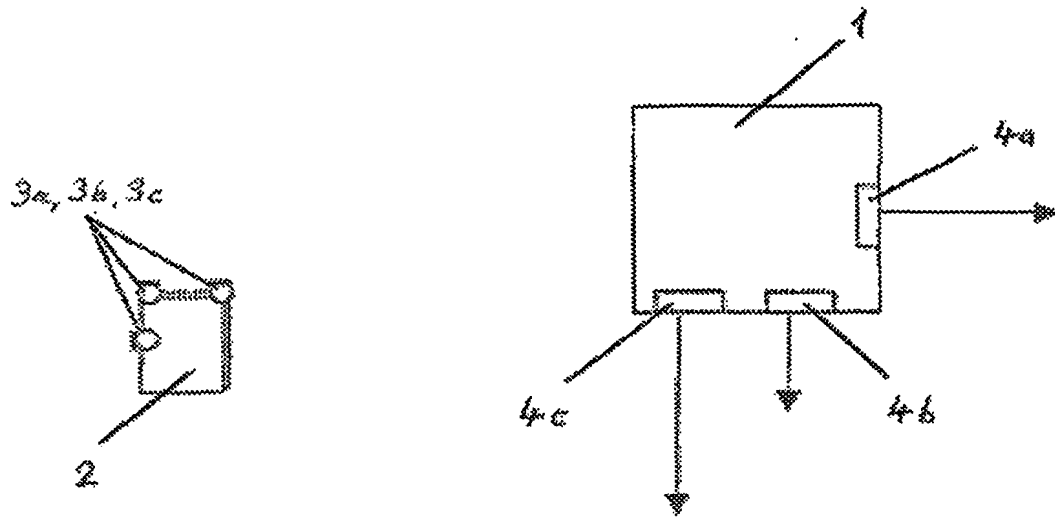


Fig. 1

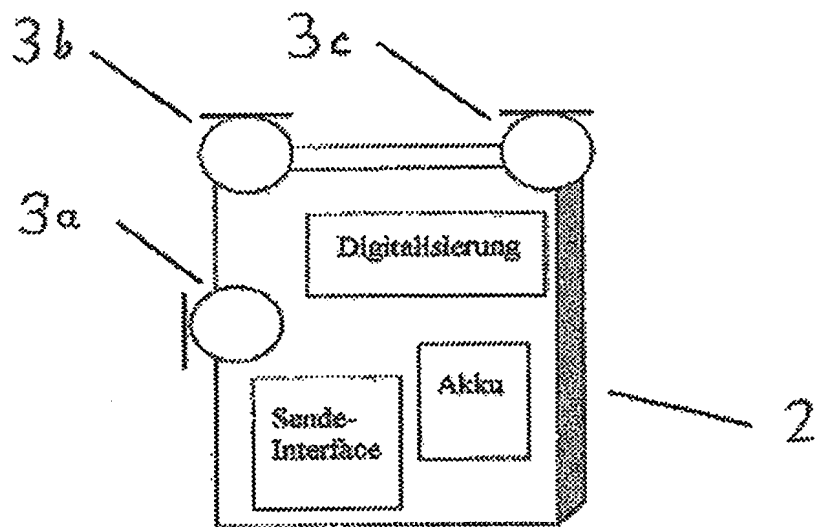
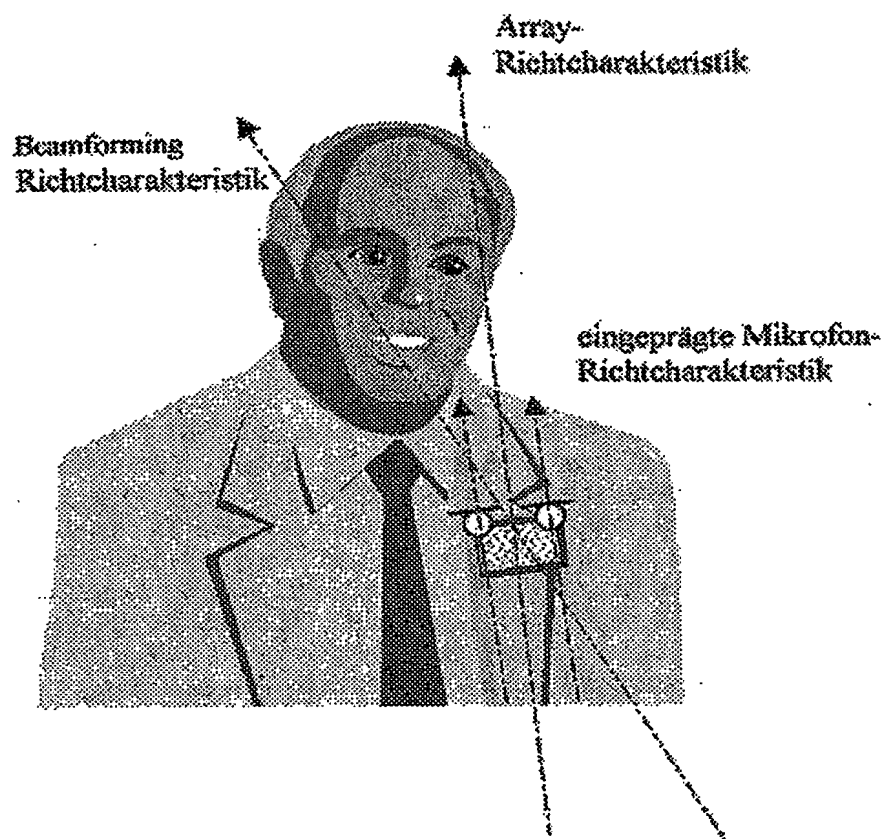


Fig. 2

5



10

Fig. 3

BEST AVAILABLE COPY